Applied and Environmental Microbiology, Aug. 2010, p. 5039–5045 Vol. 76, No. 15
0099-2240/10/$12.00 doi:10.1128/AEM.00204-10
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Hydroxyapatite-Mediated Separation of Double-Stranded DNA, Single-Stranded DNA, and RNA Genomes from Natural Viral Assemblages[▽]†

Cynthia Andrews-Pfannkoch,[2]‡ Douglas W. Fadrosh,[1]‡ Joyce Thorpe,[2] and Shannon J. Williamson[1]*

J. Craig Venter Institute, 10355 Science Center Drive, San Diego, California 92121,[1] and J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850[2]

Received 26 January 2010/Accepted 28 May 2010

Metagenomics can be used to determine the diversity of complex, often unculturable, viral communities with various nucleic acid compositions. Here, we report the use of hydroxyapatite chromatography to efficiently fractionate double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), dsRNA, and ssRNA genomes from known bacteriophages. Linker-amplified shotgun libraries were constructed to generate sequencing reads from each hydroxyapatite fraction. Greater than 90% of the reads displayed significant similarity to the expected genomes at the nucleotide level. These methods were applied to marine viruses collected from the Chesapeake Bay and the Dry Tortugas National Park. Isolated nucleic acids were fractionated using hydroxyapatite chromatography followed by linker-amplified shotgun library construction and sequencing. Taxonomic analysis demonstrated that the majority of environmental sequences, regardless of their source nucleic acid, were most similar to dsDNA viruses, reflecting the bias of viral metagenomic sequence databases.

Viruses, particularly bacteriophages (phages), are the most numerous biological entities on Earth and influence multiple biologically significant processes, from horizontal gene transfer (8) to the balance of essential nutrients in natural ecosystems (14). Viral metagenomic studies conducted over the past decade have revealed a staggering level of diversity in numerous environments and have caused a paradigm shift in our understanding of how viruses influence host physiology and evolution (2, 7). The initial discovery of photosynthesis-related genes in the genomes of cyanophages was startling (18, 26). However, this discovery has been dwarfed by the observed widespread occurrence and distribution of multiple classes of virus-encoded cellular genes in the marine environment (4, 11, 16, 17, 24–27).

The majority of viral metagenomic studies to date have focused on double-stranded DNA (dsDNA) viruses (11, 19, 27). Much less attention has been directed toward viruses with alternate genome compositions, despite their potential significance in natural ecosystems (15, 20, 22). Current library construction protocols used to study environmental DNA or RNA viruses require an initial nuclease treatment in order to remove nontargeted templates (10). Furthermore, the discrete examination of environmental single-stranded DNA (ssDNA) and RNA virus populations is complicated by the fact that traditional viral library construction methods capture only their actively replicating dsDNA forms.

This report describes the efficient fractionation and recovery of a mixture of known dsDNA, ssDNA, dsRNA, and ssRNA viral nucleic acids using hydroxyapatite (HAP) chromatography followed by linker-amplified shotgun library (LASL) construction. The fractionation of nucleic acids using HAP (a form of crystalline calcium phosphate) has been routine since the 1960s (5). This method exploits the charge interaction between positively charged Ca^{2+} ions on the surface of the HAP and the negatively charged phosphate backbone of the nucleic acids. The abundance of phosphate groups available to interact with the Ca^{2+} ions is in part dictated by the size and conformation of the nucleic acid species. Phosphate ions present in the buffer compete with the phosphate groups of the retained nucleic acid species for Ca^{2+} on the HAP. Nucleic acids with fewer available phosphate groups (e.g., circular ssDNA molecules) are not retained on the HAP as well as molecules with more available phosphate groups (e.g., dsDNA or RNA species). Differential elution of nucleic acids is typically accomplished by either the application of an increasing phosphate gradient (continuous or step) at a constant temperature or a combination of increasing temperatures and phosphate concentrations.

LASL construction has been a primary tool for the generation of Sanger sequence data from complex viral communities (7). This method leverages the ability to attach short adaptor molecules, which act as PCR primer recognition sites, to the ends of target DNA. High-quality LASLs were generated using <50 ng of input dsDNA from a known mixture of bacteriophages and sequenced using Sanger technology. These techniques were subsequently applied to nucleic acids purified from viral communities that were collected from surface aquatic waters in the Dry Tortugas National Park and a subsurface hypoxic basin within the Chesapeake Bay estuary. To the best of our knowledge, HAP chromatography has never been applied to the fractionation of nucleic acids from purified environmental viral assemblages in preparation for metagenomic sequencing.

* Corresponding author. Mailing address: J. Craig Venter Institute, 10355 Science Center Dr., San Diego, CA 92121. Phone: (858) 200-1800. Fax: (858) 200-1881. E-mail: swilliamson@jcvi.org.

‡ D.W.F. and C.A.-P. contributed equally to this work.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 11 June 2010.

TABLE 1. Metadata associated with environmental samples

Sample collection	Date collected (mo/day/yr)	Latitude	Longitude	Depth (m)	Vol (liters)	Salinity (ppt)	Temp (°C)	DO[b] (mg liter^{-1})
CBAY1[a]	7/30/2007	38°58′N	76°23′W	22	~50	19.2	24.6	0.04
CBAY2[a]	7/30/2007	38°58′N	76°23′W	22	~50	19	24.6	0.04
CBAY3[a]	7/30/2007	38°58′N	76°23′W	22	~50	18.3	24.6	0.04
CBAY4[a]	7/31/2007	38°58′N	76°23′W	22	~50	19	24.6	0.04
CBAY5[a]	7/31/2007	38°58′N	76°23′W	22	~50	NA[c]	24.6	0.04
Dry Tortugas	1/8/2004	24°29′N	83°4′W	2	~200	36	25.3	5.6

[a] Chesapeake Bay sample collections 1 to 5.
[b] DO, dissolved oxygen.
[c] NA, not applicable.

MATERIALS AND METHODS

Collection of virus-like particles from marine and estuarine environments. Approximately 200 liters of surface seawater was collected from the Dry Tortugas National Park, and five independent ~50-liter samples of estuarine water were collected from a hypoxic region of the Chesapeake Bay (Table 1). Microbial communities were size fractioned through a 20-μm Nytex mesh, followed by serial filtration through 3.0-μm (Supor), 0.8-μm (polyethersulfone), and 0.1-μm (polyethersulfone) membrane filters (Pall Life Sciences, Port Washington, NY). Virus-like particles that passed through the 0.1-μm membrane filter were concentrated via tangential flow filtration to ~1 liter using a Pellicon TFF System (Millipore, Billerica, MA) fitted with a polyethersulfone 50-kDa BioMax Maxi cassette filter. Viral concentrates were cryoprotected with molecular biology grade glycerol (10% final concentration) prior to being snap-frozen in liquid nitrogen (Chesapeake Bay) or storage onboard *Sorcerer II* at −20°C (Dry Tortugas National Park). Samples were transported to the laboratory and stored at −80°C prior to being processed.

The thawed viral concentrates were further concentrated to ~10 ml using a Centricon Plus-70 Centrifugal Filter Device (5,000 molecular weight cutoff [MWCO]; Millipore, Billerica, MA), and treated with 1 unit ml^{-1} of RNase-free DNase I for 2 h at room temperature to remove dissolved environmental DNA. The DNase reaction was terminated by the addition of EDTA and EGTA to final concentrations of 50 mM each. The DNase-treated viral concentrates were layered onto a cushion of 38% (wt/vol) sucrose in SM buffer and centrifuged at 174,899 × *g* for 3 h at 14°C using a Beckman Coulter Optima L-100 XP Ultracentrifuge equipped with a SW32Ti rotor. The pellet containing virus-like particles was dried and resuspended in SM buffer (10 mM $MgSO_4$, 100 mM NaCl, 0.01% gelatin, and 50 mM Tris-HCl). 16S rRNA PCR, using the primer set 27F/1492R (Table 2), was performed to ensure the samples were free of contaminating bacterial DNA. Nucleic acids from virus-like particles were recovered by extraction with 1 volume of phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol/vol]) and ethanol precipitation.

HAP chromatography. DNA grade Bio-Gel HTP hydroxyapatite (control number 210004557; Bio-Rad, Hercules, CA) was hydrated with 0.12 M phosphate buffer and defined according to the manufacturer's recommendations. A detailed description of the method and all buffers is provided in the supplemental material. Phosphate buffers and hydrated HAP were maintained at 60°C throughout nucleic acid fractionation. The specific nucleic acid binding and elution properties of hydrated HAP slurries can differ, so testing each batch is recommended. A sterile 0.7-cm inside diameter (i.d.) standard jacketed Econo Column equipped with a stopcock (Bio-Rad, Hercules, CA) was coated with 1 ml of Sigmacote (Sigma-Aldrich, St. Louis, MO) to prevent nonspecific nucleic acid adsorption to the column and allowed to air dry. The column was attached to a 60°C circulating water bath and rinsed twice with 7 ml of 0.12 M phosphate buffer (Fig. 1). With the stopcock closed, 2 ml of well mixed hydrated HAP was added to the column and allowed to settle for 30 min. Equilibration of the column was accomplished by washing it with 7 column volumes of 0.12 M phosphate buffer. The last wash was allowed to completely drain from the column.

A known mixture of nucleic acids consisting of M13mp18 circular ssDNA (7,249 bp), phi6 segmented linear dsRNA (6,374 bp, 4,063 bp, and 2,948 bp), MS2 linear ssRNA (3,569 bp), and lambda linear dsDNA (48,502 bp) was used to test HAP separation. Nucleic acids (known or environmental) were combined with 1/2 column volume of 0.12 M phosphate buffer. To equilibrate the sample and reduce nucleic acid secondary structure, the samples were heated to 60°C for 15 min and loaded onto the column with the stopcock closed. The nucleic acids were allowed to bind to the HAP column for 15 min. The stopcock was opened, and the initial flowthrough was collected. The column was eluted with a step gradient of 7 column volumes each of 0.12 M, 0.18 M, 0.40 M, and 1.0 M phosphate buffers for the known test mixture or 7 column volumes each of 0.12 M, 0.20 M, 0.40 M, and 1.0 M for the environmental nucleic acids. The 0.12 M fraction was pooled with the initial flowthrough. Individual fractions were extracted with equal volumes of phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol/vol]) and desalted using Amicon Ultra Centrifugal Filter Devices (30,000 MWCO; Millipore, Billerica MA) according to the manufacturer's recommendations. The fractionated nucleic acids were ethanol precipitated and resuspended in 10 μl of RNase-free 1× Tris-EDTA (TE). The fractionated nucleic acids were analyzed on a 0.8% E-gel (Invitrogen, Carlsbad, CA) poststained with 1× SYBR Gold (Invitrogen, Carlsbad, CA) and visualized using a Bio-Rad Gel Doc System (Bio-Rad, Hercules, CA). The 0.40 M and 1.0 M fractions containing dsDNA were combined.

Conversion of viral ssDNA to linear dsDNA. Purified nucleic acids from the 0.12 M HAP fractions were RNase treated to remove any carryover RNA from the HAP column. Purified circular ssDNA was converted to linear dsDNA by a pair of *Escherichia coli* DNA polymerase I reactions (Fig. 2) as described in the supplemental material. Briefly, ssDNA and random hexamers (Invitrogen, Carlsbad, CA) were combined and denatured. The hexamers were annealed to the ssDNA using positive-ramp primer annealing (4°C for 2 min, 0.1°C s^{-1} ramp to 10°C, 10°C for 10 min, 0.1°C s^{-1} ramp to 15°C, 15°C for 10 min, 0.1°C s^{-1} ramp to 25°C, and 25°C for 10 min). The hexamer-annealed ssDNA molecules were

TABLE 2. Adaptor and primer sequences

Adaptor or primer	Sequence[a]
I-CeuI adaptor molecule	5′Phos/TCGCTACCTTAGGACCGTTATAGTTA AGCGATGGAATCCTGGCAATATCAAT/5′Phos
BstXI adaptor molecule	5′Phos/CTTTCCAGCACA GAAAGGTC/5′Phos
I-CeuI bottom primer	5′Phos/TCGCTACCTTAGGACCGTTATAGTTA-3′
27F 16S rRNA forward primer	5′-AGAGTTTGATCCTGGCTCAG-3′
1492R 16S rRNA reverse primer	5′-GGTTACCTTGTTACGACTT-3′

[a] 5′Phos, 5′ phosphorylation.

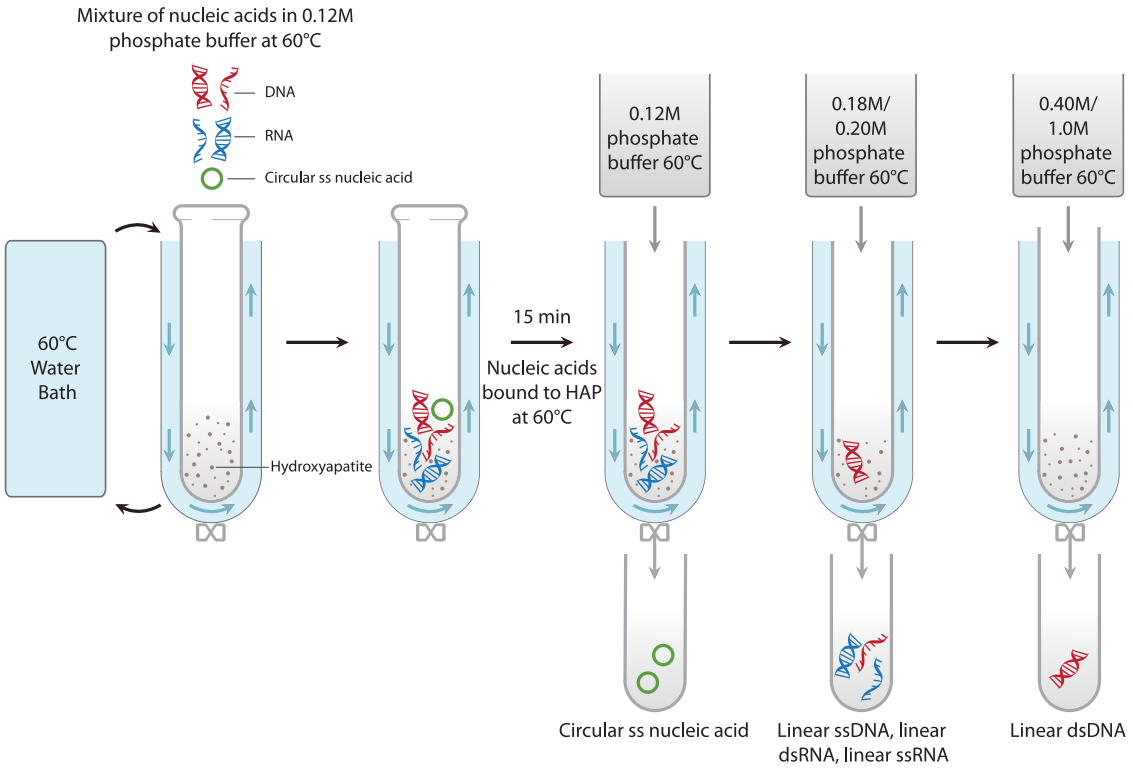

FIG. 1. Flow diagram depicting HAP-mediated separation of known and environmental viral nucleic acids.

reacted with *E. coli* DNA polymerase I at 16°C for 2 h. The reaction mixture was extracted with 1 volume of phenol-chloroform-isoamyl alcohol, ethanol precipitated, and resuspended in RNase-free 1× TE. This process was repeated using the product from the first reaction as the input for the second reaction.

Conversion of viral RNA to linear dsDNA. The 0.18 M (known mixture) and the 0.20 M (environmental) HAP fractions were DNase I treated to remove possible carryover DNA contamination from the HAP column. Purified RNA was converted to dsDNA using the Superscript RNA Amplification System Kit (Invitrogen, Carlsbad, CA) with minor modifications. Random hexamers (50 ng) were mixed with the HAP-purified RNA in 15 μl of diethyl pyrocarbonate (DEPC)-treated water, heated to 70°C for 15 min, and placed directly on ice. The hexamers were annealed to the RNA using the same positive-ramp primer annealing described above. A first-strand synthesis reaction mixture was prepared as described by the manufacturer and incubated at 50°C for 1 h, 55°C for 1 h, and 70°C for 15 min. The first-strand synthesis was added to the second-strand synthesis reaction, which was performed according to the manufacturer's recommendations. The reaction mixture was extracted with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1 [vol/vol/vol]) and ethanol precipitated. The pellet was resuspended in 20 μl RNase-free 1× TE.

Linker-amplified shotgun library construction. LASLs were constructed from the dsDNA forms from each of the HAP fractions as described by Adams and colleagues (1) with minor modifications (a detailed description of the method is

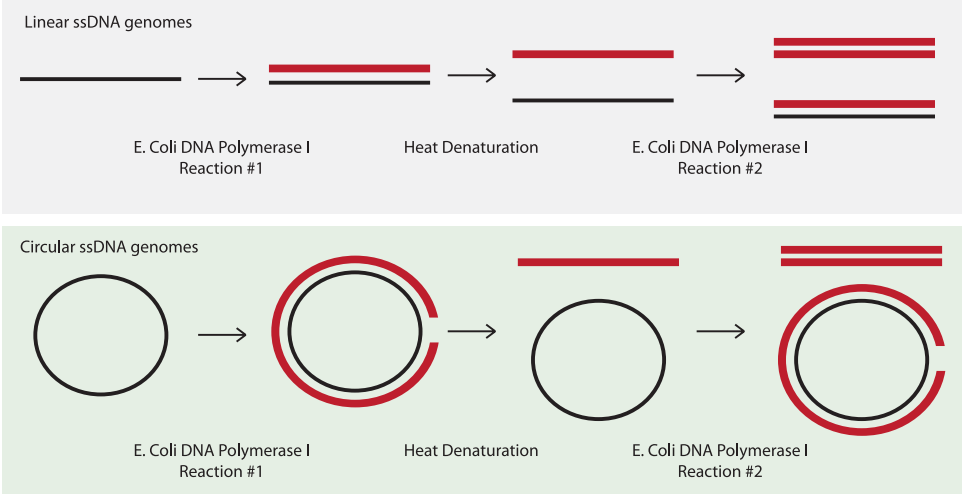

FIG. 2. Conversion of ssDNA genomes to linear dsDNA. Single-stranded DNA templates are converted to linear dsDNA molecules by sequential treatment with *E. coli* DNA polymerase I in the presence of random hexamers.

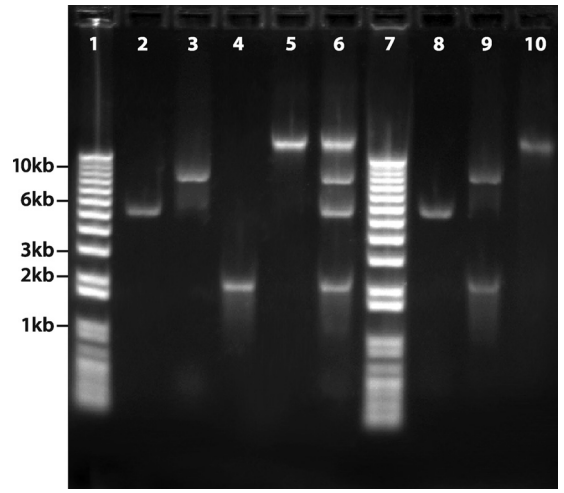

FIG. 3. Separation of known viral nucleic acids using HAP chromatography. Ten nanograms of M13mp18 ssDNA (lane 2), phi6 dsRNA (lane 3), MS2 ssRNA (lane 4), and lambda dsDNA (lane 5), respectively, were combined (lane 6) and visualized, along with 1-kb Plus DNA Ladder (lanes 1 and 7) on a 0.8% E-gel. One hundred nanograms of each nucleic acid were combined, loaded onto a HAP column, and eluted using 0.12 M (lane 8), 0.18 M (lane 9), and 0.40 M/1.0 M (lane 10) phosphate buffer. Lanes 8 to 10 show 1/10th (~10 ng) of the HAP-purified nucleic acids.

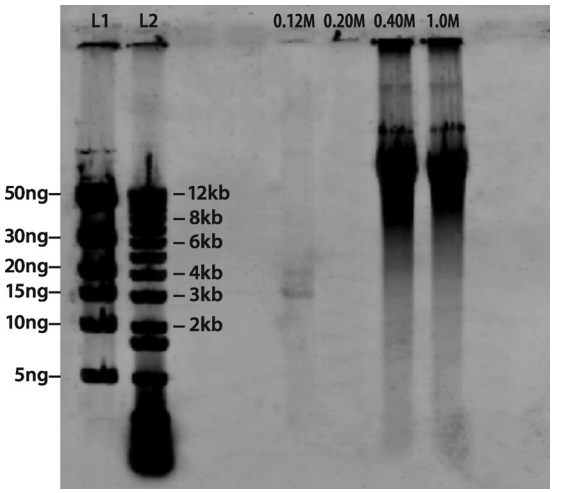

FIG. 4. Separation of viral-community nucleic acids from a subsurface hypoxic region of the Chesapeake Bay. Viral nucleic acids purified from the Chesapeake Bay (Collection CBAY1 is shown) were fractionated on a HAP column using 0.12 M, 0.20 M, 0.40 M, and 1.0 M phosphate buffer and visualized with 1 μl of High Mass DNA Ladder (L1) and 1-kb Plus DNA Ladder (L2).

provided in the supplemental material). Briefly, the dsDNAs were randomly sheared, end polished with Bal31 nuclease-T4 DNA polymerase, and carefully size selected on 1% low-melting-point agarose. I-CeuI adaptors (Table 2) (Integrated DNA Technologies, Coralville, IA) were appended to the ends of the DNA fragments, followed by amplification with Phusion DNA polymerase (Finnzymes, Espoo, Finland) for 15 cycles. The ends of the amplicons were polished, and BstXI adaptors (Table 2) were appended to the fragments. The fragments were then inserted into a BstXI-linearized medium-copy plasmid vector containing transcriptional terminators flanking the cloning sites. The transformed cells were processed and sequenced using the high-throughput sequencing pipeline at the J. Craig Venter Institute (Rockville, MD) as described by Rusch et al. (23). Sequences from the mixture of known viruses and environmental virus sequences were deposited in NCBI's Trace Archive (see the supplemental material for Trace Archive identifiers).

Phylogenetic analysis of environmental sequences. The taxonomic affiliations of environmental virus sequences were determined using the Automated Phylogenetic Inference System (APIS) (3, 6). Open reading frames (ORFs) were predicted using MetaGene (21), and query protein sequences were compared to sequences within APIS databases that contained (i) all publicly available microbial genomes, including bacteria, archaea, and single-celled eukaryotes; (ii) all fully sequenced viral genomes deposited in the NCBI's Viral Genomes Resources; and (iii) all publicly available Sanger-based "long-read" viral metagenomes. Homologs to the query sequences were identified, and their full-length sequences were aligned to query proteins using MUSCLE (12). Neighbor-joining phylogenetic trees were produced, and bootstrap values were calculated based on 100 replicates. The taxonomy of the query sequence was inferred from its placement on the tree. Three homologs were required for tree placement.

RESULTS AND DISCUSSION

Validation of viral nucleic acid fractionation by HAP chromatography. Viral nucleic acids applied to a HAP column were eluted with a discontinuous step gradient of increasing concentrations of phosphate-containing buffer (Fig. 1). As anticipated, the viruses in the test mixture eluted at phosphate buffer concentrations of 0.12 M (M13mp18 circular ssDNA), 0.18 M (phi6 segmented linear dsRNA and MS2 linear ssRNA), and 0.40/1.0 M (lambda linear dsDNA). Nucleic acids with more phosphate on the backbone eluted at higher phosphate concentrations. Comparison of input and eluted nucleic acids via gel electrophoresis indicated that all were recovered at their expected molecular weights (Fig. 3). This observation suggests that HAP chromatography is a highly efficient method for the fractionation and recovery of known viral nucleic acids.

HAP-fractionated nucleic acids purified from the Chesapeake Bay (Fig. 4) and the Dry Tortugas National Park (data not shown) were visualized via gel electrophoresis. Environmental viral nucleic acid species exhibited an elution profile similar to that of the test mixture and migrated within a molecular weight range consistent with known environmental viruses of similar nucleic acid compositions. This validated our ability to fractionate environmental viral nucleic acids efficiently using HAP chromatography.

Conversion of circular ssDNA to linear dsDNA. Single-stranded DNA molecules are converted to linear dsDNA molecules using a pair of sequential *E. coli* DNA polymerase I reactions in the presence of random hexamers. The 5′-3′ exonuclease activity of the polymerase nick translates upstream of the newly synthesizing strand, producing a linear single-stranded copy of ssDNA templates. The second reaction synthesizes a DNA strand complementary to the linear ssDNA copy created in the first reaction. This effectively yields a linear duplex copy of linear and circular ssDNA genomes (Fig. 2). Even though this mechanism skews community diversity by producing a pair of duplex molecules for every linear single-stranded template, it is vital for the generation of duplex forms of circular templates, since an efficient method to randomly shear small circular molecules is not available.

Linker-amplified shotgun library construction. Linear dsDNAs were randomly sheared to interrupt potentially disruptive DNA sequences and to generate fragment sizes suitable for robust PCR amplification and cloning into *E. coli*. Amplification of I-CeuI-terminated fragments was restricted to 15 cycles using Phusion DNA polymerase to minimize the occurrence of

TABLE 3. Sequencing analysis of LASLs constructed from known viral genomes

Library	No. of clones/library	Total reads	No. of nontrash reads (% success)	Avg length (bp)	Composition (%)			
					M13mp18[a]	MS2[b]/phi6[c]	Lambda[d]	Other
0.12 M	$\sim 5.70 \times 10^5$	96	44 (45.8)	756	90.90	6.80	2.30	0.00
0.18 M	$\sim 1.75 \times 10^6$	96	86 (89.6)	770	0.00	100.00	0.00	0.00
0.4/1.0 M	$\sim 1.38 \times 10^6$	96	78 (81.3)	760	0.00	1.30	94.80	3.90

[a] Bacteriophage M13mp18 contains a 7,249-bp circular ssDNA genome.
[b] Bacteriophage MS2 contains a 3,569-bp linear ssRNA genome.
[c] *Pseudomonas* phage phi6 contains a segmented linear dsRNA genome. Segments L, M, and S have 6,374 bp, 4,063 bp, and 2,948 bp respectively.
[d] Bacteriophage lambda contains a 48,502-bp linear dsDNA genome.

random mutations, chimeras, and bias, at the same time producing sufficient quantities of DNA for library construction. The plasmid vector contained transcriptional terminators flanking the insertion sites to prevent transcription of potentially lethal insert sequences. The transformed plasmids were analyzed, and a significant proportion were found to contain inserts of the targeted size. Transformation titers were suitable for typical Sanger sequencing projects, and the overall library quality was assessed through the analysis of sequencing-data results (Tables 3 and 4).

A direct comparison of Sanger sequencing reads produced from linker-amplified and unamplified marine viral-community DNA demonstrated no significant changes in the relative abundances of viral genotypes based on BLAST analysis (unpublished data). This suggests that the LASL approach does not introduce significant biases that could potentially influence the interpretation of the environmental virus community data.

Validation of separation efficiency through sequence analysis. The taxonomic analysis of libraries created from known viral nucleic acids using BLASTn revealed that at least 90% of the sequencing reads generated from the purified, HAP-fractionated nucleic acids had significant similarity to the expected viral genomes (Table 3). APIS, an alternative taxonomic analysis tool to BLAST, was used in the analysis of the environmental libraries. APIS is a more robust tool than BLAST for examining the taxonomy of environmental metagenomic sequences, since it utilizes phylogenetic relationships between query and full-length reference sequences in order to infer taxonomy (Fig. 5). This method is superior to BLAST-based analyses, since the "top-BLAST" hit for many viral metagenomic sequences is often uninformative (e.g., hypothetical or with similarity to cellular genes).

Predicted environmental protein sequences that were characterized using APIS and claded with viruses were subjected to further scrutiny. Between 57% and 91% of sequences were not placed on phylogenetic trees by APIS and remained uncharacterized. The highest proportion of sequences from every environmental library, regardless of initial nucleic acid composition, claded with dsDNA viruses despite the fact that ssDNA- and RNA-specific libraries were constructed. In addition, an average of 67% of ssDNA and 17% of RNA sequences claded with dsDNA viruses from marine metagenomes (see Table S1 in the supplemental material). These findings are consistent with two observations: (i) the majority of publicly available viral metagenomes represent dsDNA viruses, and (ii) dsDNA viruses comprise only ~35% of the fully sequenced viral genomes within the NCBI Viral Genomes Resource, yet they represent the largest total proportion of genetic material available for comparison on a nucleotide-per-nucleotide basis.

Approximately 15% and 50% of the sequences generated from the HAP-fractionated dsDNA from the Chesapeake Bay and the Dry Tortugas, respectively, claded with dsDNA viruses, with a limited number sharing similarity with ssDNA or RNA viral genomes (Table 4). Even though sequences generated from the HAP-fractionated environmental ssDNA and RNA were most similar to dsDNA viruses, they also shared

TABLE 4. Sequence analysis of LASL libraries from environmental viral communities

Library	No. of clones/library	Total reads	No. of nontrash reads (% success)	Avg length (bp)	ssDNA virus related		RNA virus related		dsDNA virus related	
					No. of predicted proteins	% Predicted proteins[c]	No. of predicted proteins	% Predicted proteins[c]	No. of predicted proteins	% Predicted proteins[c]
Chesapeake Bay hypoxic basin										
ssDNA[a]	$\sim 9.00 \times 10^5$	6,144	5,786 (92.4)	776	277	23.01	0	0.00	830	68.94
RNA[a]	$\sim 7.10 \times 10^5$	6,144	5,787 (94.2)	758	158	16.83	2	0.21	189	20.13
dsDNA[b]	$\sim 1.88 \times 10^6$	95,615	87,900 (91.6)	734	0	0.00	0	0.00	1,076	14.79
Dry Tortugas National Park										
ssDNA	$\sim 1.91 \times 10^5$	6,144	5,803 (94.4)	745	2	0.13	0	0.00	444	29.38
RNA	$\sim 3.37 \times 10^4$	6,144	5,589 (91.0)	780	0	0.00	1	0.03	225	6.28
dsDNA	$\sim 1.28 \times 10^6$	90,624	68,491 (75.6)	755	0	0.00	39	0.57	3,445	50.06

[a] Nucleic acids from five independent sample collections were pooled, and a single library was constructed.
[b] Data were compiled from four libraries constructed from four independent sample collections.
[c] Percentage of total predicted proteins (those placed on trees by APIS) that clade with the specified genome type. Nonviral classifications are not reported.

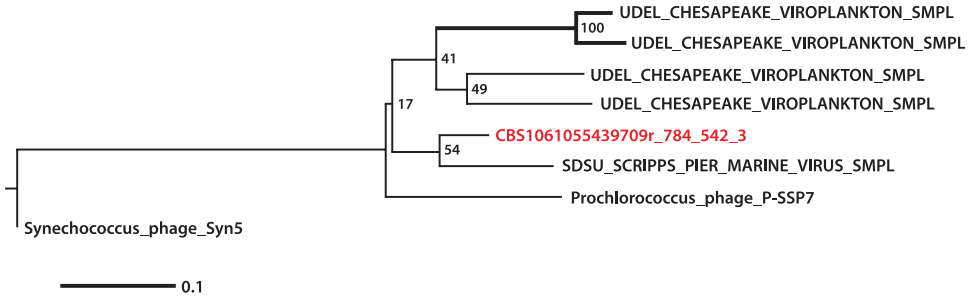

FIG. 5. Neighbor-joining phylogenetic tree produced by APIS. An example of a Chesapeake Bay ssDNA viral query sequence (depicted in red) clading with sequences from other dsDNA environmental viral metagenomes. The viral metagenome called UDEL_CHESAPEAKE_VIRIOPLANKTON_SMPL was generated from a separate study (3). The viral metagenome called SDSU_SCRIPPS_PIER_MARINE_VIRUS_SMPL was generated by a study conducted by Breitbart et al. (7). Bootstrap values are indicated at branches. The scale bar indicates 0.1 amino acid substitution per position.

similarity with known ssDNA- and RNA-containing viruses. For the Chesapeake Bay ssDNA library, 23% of characterized sequences were most similar to known ssDNA viruses. A very small percentage (0.21%) of sequences from the Chesapeake Bay RNA library were most similar to known RNA viruses.

The taxonomic distribution of sequences generated from the Dry Tortugas ssDNA and RNA libraries differed drastically from those constructed from the Chesapeake Bay, suggesting that inherent differences in viral community composition exist. Only 0.13% of APIS-characterized sequences from the ssDNA library and 0.03% of sequences from the RNA library claded with other known ssDNA and RNA viruses, respectively. Although the inferred taxonomy of the viral sequences from the environmental libraries does not directly correlate with viruses of the expected genome composition, the results of the fractionation experiment using known nucleic acids clearly demonstrated that efficient HAP-mediated separation of viral genotypes does occur.

Benefits of HAP chromatography and LASL construction for viral ecology. Nucleic acid fractionation by HAP chromatography and subsequent LASL construction has marked benefits over current purification and library construction protocols. HAP chromatography eliminates the need to choose between studying either DNA or RNA viruses while allowing the direct targeting and isolation of all viral nucleic acid types from environmental samples.

Converting non-dsDNA viral genomes to a linear duplex form standardizes library construction efforts and facilitates the comparison of sequencing results across HAP fractions, since the same methodologies can be used for all nucleic acid species. Having a dedicated library construction pipeline that can accommodate all nucleic acid types reduces potential biases associated with nucleic acid-specific library construction methods (9, 10, 15, 20). The creation of high-quality viral metagenomic libraries from as little as 2 ng of starting nucleic acid (data not shown) using LASL construction is more reliable than alternative amplification methods (e.g., multiple-displacement amplification [MDA]).

Several known challenges are associated with MDA of mixed microbial communities. They include bias toward circular ssDNA templates, the production of chimeras, and nonspecific amplification, which can be mistaken for novel viral sequence during downstream analyses. Linker amplification of viral-community nucleic acid is a robust method for generating quantities of DNA required for metagenomic sequencing using Sanger technology. Furthermore, the LASL construction method has been easily adapted to accommodate next-generation sequencing platforms (e.g., 454 pyrosequencing) since the linker amplification PCR step produces sufficient quantities of DNA for this application.

The interpretation of viral metagenomic data from DNA or RNA can be cumbersome for reasons including genetic novelty (e.g., no shared homology to known sequences) (13), the abundance of virus-encoded cellular genes (11, 27), and the lack of suitable environmental virus reference genomes available for comparison. Furthermore, certain members of the viral community will be retained on larger prefilters, which may influence interpretation of the overall diversity of viral communities. HAP chromatography provides a rigorous mechanism for partitioning environmental viral nucleic acids, ensuring that the metagenomic sequences associated with each fraction accurately represent the targeted nucleic acid type. Comparison of metagenomic data across nucleic acid types enables a more complete analysis of the gene complement of viral communities and evaluation of virus diversity. Specific questions regarding the distribution of gene families across viruses with differing nucleic acid compositions and the mobilization of genes between these viruses can also be addressed using the methods described here. Lastly, HAP chromatography followed by LASL construction can be a powerful tool for the discovery of novel dsDNA, ssDNA, dsRNA, and ssRNA viruses in the environment.

ACKNOWLEDGMENTS

This research was supported by the Office of Science (BER), U.S. Department of Energy, Cooperative Agreement no. De-FC02-02ER63453, and by the National Science Foundation's Microbial Genome Sequencing Program (award number 0626826).

We thank John Glass for his technical expertise and advice, Lisa Zeigler for her assistance with APIS analyses, and K. Eric Wommack for his assistance with environmental sample collection.

REFERENCES

1. **Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G.**

Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, et al. 2000. The genome sequence of Drosophila melanogaster. Science **287:**2185–2195.

2. **Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer.** 2006. The marine viromes of four oceanic regions. PLoS Biol. **4:**e368.
3. **Badger, J. H., T. R. Hoover, Y. V. Brun, R. M. Weiner, M. T. Laub, G. Alexandre, J. Mrazek, Q. Ren, I. T. Paulsen, K. E. Nelson, H. M. Khouri, D. Radune, J. Sosa, R. J. Dodson, S. A. Sullivan, M. J. Rosovitz, R. Madupu, L. M. Brinkac, A. S. Durkin, S. C. Daugherty, S. P. Kothari, M. G. Giglio, L. Zhou, D. H. Haft, J. D. Selengut, T. M. Davidsen, Q. Yang, N. Zafar, and N. L. Ward.** 2006. Comparative genomic evidence for a close relationship between the dimorphic prosthecate bacteria Hyphomonas neptunium and Caulobacter crescentus. J. Bacteriol. **188:**6841–6850.
4. **Bench, S. R., T. E. Hanson, K. E. Williamson, D. Gosh, M. Radosovich, K. Wang, and K. E. Wommack.** 2007. Metagenomic characterization of Chesapeake Bay virioplankton. Appl. Environ. Microbiol. **73:**7629–7641.
5. **Bernardi, G.** 1965. Chromatography of nucleic acids on hydroxyapatite. Nature **206:**779–783.
6. **Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J. P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kroger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jezequel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M. P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J. L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, and I. V. Grigoriev.** 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456:**239–244.
7. **Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer.** 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U. S. A. **99:**14250–14255.
8. **Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brussow.** 2003. Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. **6:**417–424.
9. **Culley, A. I., A. S. Lang, and C. A. Suttle.** 2003. High diversity of unknown picorna-like viruses in the sea. Nature **424:**1054–1057.
10. **Culley, A. I., A. S. Lang, and C. A. Suttle.** 2006. Metagenomic analysis of coastal RNA virus communities. Science **312:**1795–1798.
11. **Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer.** 2008. Functional metagenomic profiling of nine biomes. Nature **455:**830.
12. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.
13. **Edwards, R. A., and F. Rohwer.** 2005. Viral metagenomics. Nat. Rev. Microbiol. **3:**504–510.
14. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. Nature **399:**541–548.
15. **Lang, A. S., M. L. Rise, A. I. Culley, and G. F. Steward.** 2009. RNA viruses in the sea. FEMS Microbiol. Rev. **33:**295–323.
16. **Lindell, D., M. B. Sullivan, Z. I. Johnson, A. C. Tolonen, F. Rohwer, and S. W. Chisholm.** 2004. Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc. Natl. Acad. Sci. U. S. A. **101:**11013–11018.
17. **Mann, N. H., M. R. J. Clokie, A. Millard, A. Cook, W. H. Wilson, P. J. Wheatley, A. Letarov, and H. M. Krisch.** 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. J. Bacteriol. **187:**3188–3200.
18. **Mann, N. H., A. Cook, A. Millard, S. Bailey, and M. Clokie.** 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature **424:**741.
19. **McDaniel, L., M. Breitbart, J. Mobberley, A. Long, M. Haynes, F. Rohwer, and J. H. Paul.** 2008. Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. PLoS One **3:**e3263.
20. **Ng, T. F. F., C. Manire, K. Borrowman, T. Langer, L. Ehrhart, and M. Breitbart.** 2009. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. J. Virol. **83:**2500–2509.
21. **Noguchi, H., J. Park, and T. Takagi.** 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res. **34:**5623–5630.
22. **Rosario, K., S. Duffy, and M. Breitbart.** 2009. Diverse circovirus-like genome architectures revealed by environmental metagenomics. J. Gen. Virol. **90:** 2418–2424.
23. **Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. **5:**e77.
24. **Sharon, I., A. Alperovitch, F. Rohwer, M. Haynes, F. Glaser, N. Atamna-Ismaeel, R. Y. Pinter, F. Partensky, E. V. Koonin, Y. I. Wolf, N. Nelson, and O. Beja.** 2009. Photosystem I gene cassettes are present in marine virus genomes. Nature **461:**258–262.
25. **Sharon, I., S. Tzahor, S. Williamson, M. Shmoish, D. Man-Aharonovich, D. B. Rusch, S. Yooseph, G. Zeidner, S. S. Golden, S. R. Mackey, N. Adir, U. Weingart, D. Horn, J. C. Venter, Y. Mandel-Gutfreund, and O. Beja.** 2007. Viral photosynthetic reaction center genes and transcripts in the marine environment. ISME J. **1:**492–501.
26. **Sullivan, M. B., M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm.** 2005. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol. **3:**e144.
27. **Williamson, S. J., D. B. Rusch, S. Yooseph, A. L. Halpern, K. B. Heidelberg, J. I. Glass, C. Andrews-Pfannkoch, D. Fadrosh, C. S. Miller, G. Sutton, M. Frazier, and J. C. Venter.** 2008. The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS One **3:**e1456.